A New Approach to Automatic Term Extraction

Ádám Kis  Balázs Kis  Gábor Pohl
SZAK Publishers Ltd.  MorphoLogic Ltd.  Pázmány Péter Catholic University,
adam.kis@szak.hu  kis@morphologic.hu  Fac. of Information Technology

An ideal term extraction system is capable of finding terms in previously unknown source texts without human intervention, and with large recall and precision. However, term occurrences have semantic, syntactic and discourse-related characteristics, so this task raises a modelling problem common to all fields of computational linguistics: for the sake of efficiency and feasibility, most linguistic phenomena must be assigned one or more surface characteristics.

This paper starts with emphasizing some definition problems related to term extraction. Then the authors describe a project aiming at the development of a term extraction system using a new approach. The model they employ here traces back the problem of term recognition to two basic attributes of lexemes: their terminological position and terminological role. The paper addresses the representation of nets of terms, and calls attention to the language- and topic-dependent nature of terminology.

The authors conclude the paper by comparing the present approach to those described in literature, and possible evaluation procedures.